Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

Multi-pronged abundance prediction of bee pests' spatial proliferation in Kenya

David Masereti Makori [a,b,*], Elfatih M. Abdel-Rahman [a,b], John Odindi [b], Onisimo Mutanga [b], Tobias Landmann [a], Henri E.Z. Tonnang [a,b]

[a] *International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772, Nairobi 00100, Kenya*
[b] *School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa*

A B S T R A C T

Bee farming and beehealth are threatened by climate change, agricultural and agrochemicals intensification, and bee pests and diseases. Among these threats, bee pests have particularly been identified as a major obstacle to beehealth. Although previous studies have endeavoured to establish bee pests' spatial distribution, their seasonal abundance in the landscape remains poorly understood. Hence, this study sought to determine factors that influence the abundance and spatial proliferation of bee pests in Kenya. Abundance data on *Varroa destructor, Oplostomus haroldi, Galleria mellonella* and *Aethina tumida* were collected from apiaries in Kenya during the wet and dry seasons. The abundance data were fitted to non-conflating human footprint datasets, satellite derived vegetation phenological, topographical and bioclimatic variables. The results indicated a significant ($p \leq 0.05$) seasonal influence on bee pests' abundance, while precipitation was the most relevant on most bee pests' abundance prediction models. Topographic and vegetation phenological influence varied across the landscapes while anthropogenic influence was comparatively low. High seasonality in bioclimatic variables influenced the projected (year 2055) spatial and abundance risk levels of bee pests across the study area. The *V. destructor* and *A. tumida* prediction models for current and future epochs ranked excellent in their performance, while *O. haroldi* and *G. mellonella* were ranked good and fair, respectively. Due to their precision, this study concluded that these models could reliably be used to establish bee pests' high-risk areas for management and mitigation purposes.

1. Introduction

Bee pollination services are vital for biodiversity, crop production, improving food and nutritional security, and conservation of bio(geo) diversity (Katumo et al., 2022). In semi-arid Africa with erratic rainfall which is unable to support rainfed agriculture, bee farmers directly benefit from honey and other hive by-products. Bee farming provides supplemental income from hive products such as honey and royal jelly (Raina et al., 2011) and enhance conservation efforts. Nevertheless, bee farming is threatened by climate change, agricultural and agrochemical intensification, habitat alterations, and bee pests and diseases (Muli et al., 2014).

Bee pests are particularly devastating through colony collapse due to direct physical injury or indirectly as vectors of pathogens that transmit diseases. The impact caused by pests on bee colonies has recently attracted profound scientific interest on performance (Fombong et al., 2012; Muli et al., 2014), spatial extents and proliferation patterns (Adan

et al., 2021; Kganyago et al., 2018; Makori et al., 2017), and their economic impact (Boncristiani et al., 2021). Globally, the most common and economically important bee pests are the varroa mites, large hive beetles, small hive beetles and wax moths (Fombong et al., 2012; Pirk et al., 2016; Torto et al., 2010). Although these pests are known to proliferate across agroecological gradients, their abundance, spatial and temporal distribution in Kenya remain largely unknown.

Bee pests have different optimal bioclimatic conditions with significant variations in temperature, precipitation, net productivity and altitudinal range (Peterson and Nakazawa, 2008). The habitats' biotic conditions such as net productivity and vegetation phenology affect bee pests' population dynamics and richness. Specifically, biotic conditions affect pest hosts' agility, vigour and ability to produce and accumulate hive products such as bee bread, on which the pests thrive (Baumann et al., 2017; Pirk et al., 2016). On the other hand, bioclimatic conditions could limit or enhance bee pests' proliferation and reproduction (Torto et al., 2010). Moreover, anthropogenic effects on bee pests and their

---

hosts influence their spatial distribution and proliferation (Winfree et al., 2009). For instance, landscape fragmentation and agricultural intensification adversely influence landscape continuity, and flower diversity and intensity, sustaining the pests' proliferation (Boncristiani et al., 2021). Conversely, human settlements impact bee distribution by reducing their nesting sites, habitats and their ability to resist pests' invasion. Therefore, information on human effects on the distribution of both bees and their pests are desirable.

To understand and establish the spatial distribution and proliferation of bee pests, machine learning (ML) algorithms provide statistical pathways for linking a feature(s) of interest (response) to reliable predictor variables such as biotic, bioclimatic and anthropogenic factors in space and time (Diao and Wang, 2014; Strebel et al., 2022; Uusitalo et al., 2019; Wulder et al., 2022). To achieve accurate and realistic bee pests' predictions, predictor variables should be carefully and meticulously selected. Satellite-based variables such as vegetation phenology and derived human footprint are continuous observations with better spatial and temporal resolutions that could capture 'actual' landscape patterns and human distribution patterns (Cord et al., 2014; WCS and CIESIN, 2005; Zhang et al., 2013). Moreover, response variables such as bee pests' abundance, with the ability to provide extra information other than 'location-only' are necessary for developing optimum management options.

Studies like Makori et al. (2017) have predicted the spatial distribution of bee pests using presence-only data, climate and vegetation variables. As previously mentioned, a number of studies have shown the value of abundance data in providing reliable information (Baldridge et al., 2016; Dallas and Hastings, 2018; Hallman and Robinson, 2020; Waldock et al., 2022). Moreover, Makori et al. (2017) used the maximum entropy (MaxEnt) model to predict bee pests' distribution using presence-only dataset while there are other algorithms that could handle quantitative observations such as abundance, which could be explored. Other cutting-edge machine learning algorithms such as random forest (RF) (Breiman, 2001) could perform better than MaxEnt in predicting the abundance and distribution of species. The RF algorithm has been demonstrated to yield high prediction precision and accuracy among regression methods (Belgiu and Drăguţ, 2016; Izquierdo-Verdiguier and Zurita-Milla, 2020; Parmar et al., 2019; Speiser et al., 2019). Furthermore, the influence of human footprint has not yet been included in predicting abundance distribution of bee pests in Kenya. This study hypothesized that human footprint could adequately explain anthropogenic influence on the distribution of bee pests in Kenya. Therefore, this study sought to establish the influence of bioclimatic, vegetation phenology, topography, and human footprint variables on the current and future spatial and abundance proliferation of bee pests in Kenya.

2. Methods

2.1. Study site

The study was conducted in 38 counties in Kenya encompassing four

Fig. 1. Study area on the human footprint (human influence index (HII) backdrop indicating regions used to predict proliferation of bee pests in Kenya (within the blue outline) and bee pest abundance locations (black points). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

agroecological zones. The study area covers 253,627.10 km^2 and spans more than half of the country including Nyanza, Western, Rift Valley, Mount (Mt.) Kenya, Eastern and Coastal regions (Fig. 1). Abundance data were collected from 14 counties comprising of Vihiga and Kakamega (in the western region), Nandi and Baringo (Rift Valley region), Embu, Nyeri, Laikipia, Meru and Isiolo (Mt. Kenya region), Machakos and Kitui (Eastern region) and Taita Taveta, Kilifi and Kwale (coastal region). The study sites varied in elevation from 2 m above sea level at the coastal region to 2,045 m above sea level at the Nandi Hills in the Rift Valley region.

The coastal region is a low altitude area with high temperature and humidity, with Arabuko Sokoke as the major forest. Human settlement and intensification of agriculture are two major threats to natural vegetation in the region (Schürmann et al., 2020). Mt. Kenya, Nandi and Western regions are characterized by high altitude, low temperature and high humidity. Major water towers such as Mau Complex, Cherangani Hills and Kakamega Forest are found in these regions. The Tana River County in the coastal region, Baringo County in the Rift Valley region and the larger Eastern region are characterized by high temperature and low humidity, and are predominantly covered by shrubs and few stands of Acacia trees. Due to low precipitation, small scale subsistence farming is the major socio-economic activity practiced in these regions (Schürmann et al., 2020). On the other hand, Kakamega, Kisii Highlands, Nandi Hills and Mt. Kenya regions receive higher precipitation, hence higher agricultural productivity and are mainly utilized for maize farming, a staple food crop in Kenya. In addition, bee farming is used to improve household income from hive products and crop production through pollination services (Warui et al., 2018).

2.2. Data collection and pre-processing

Data used for prediction of bee pests' proliferation were categorized into response and predictor variables.

2.2.1. Response variables

Bee pests' abundance data were collected between March and April 2014 and June 2015 wet and dry seasons, respectively. A sample point consisted of an apiary with five or more hives in one location. The apiaries were randomly selected for sampling with help from the local community according to the requirements of estimation density across the landscape in proportion to population density (Endo et al., 2015). An enumeration of all bee pests in selected apiaries was done and data on four bee pests species with global economic importance were recorded; i.e., varroa mites (*Varroa destructor*), large hive beetles (*Oplostomus haroldi*), small hive beetles (*Galleria mellonella*) and wax moths (*Aethina tumida*), using standard methods as detailed in Dietemann et al. (2013) and Torto et al. (2010). A total of 45 apiaries had *V. destructor*, 24 had *O. haroldi*, 38 had *G. mellonella* and 37 had *A. tumida* (refer to Table 2 for bee pest abundance data), which were acceptable sample size ranges for most machine learning modelling algorithm applications (Peterson and Nakazawa, 2008).

The apiaries considered for the four bee pests span across four agroecological zones in the study area, with a representative climatic gradient. The Wilcoxon rank sum test was performed to test the significance influence ($p \leq 0.05$) of seasonality on abundance of bee pests between wet and dry seasons, following heterogeneous variance observations (Fagerland and Sandvik, 2009). Further, this dataset was partitioned to 70% for training and 30% for validation purposes before the models were developed.

2.2.2. Predictor variables

All predictor variables were clipped to the boundaries of the study area and resampled and harmonized to the same (base) spatial resolution as detailed by Makori et al. (2022). Therefore, the align raster option in QGIS software (QGIS Development Team, 2022) was used to edit the resolution of topographical (90 m), bioclimatic (1 km) and human

footprint (1 km) datasets to that of the vegetation phenology datasets that had moderate resolution (250 m).

2.2.2.1. Vegetation phenology. Vegetation phenology was derived from enhanced vegetation index (EVI), acquired by a 250 m resolution Moderate Resolution Imaging Spectroradiometer (MODIS) imagery at a 16-day interval. This study used EVI observations from 2000 to 2021 (21 years) to derive vegetation phenology variables in TIMESAT software environment (Eklundha and Jönsson, 2017). Best fitting was achieved using TIMESAT parameter settings as recommended by Makori et al. (2017).

A total of 13 vegetation phenological variables were derived and used in this study. These were start of the season time (*start_t*), end of season time (*end_t*), length of season (*length*), base value (*base*), time for middle of season (*mid*), maximum value (max), amplitude (*amplitude*), left derivative (*left_d*), right derivative (*right_d*), large integral (*large_i*), small integral (*small_i*), start of season value (*start_v*) and end of season value (*end_v*).

2.2.2.2. Topographical variables. Topographical variables derived from a 90-metre pixel resolution (3 arcsec resolution) digital elevation model (DEM), were used to model the influence of land morphology on spatial distribution and proliferation of bee pests in Kenya. The DEM was acquired by a Shuttle Radar Topography Mission (SRTM) instrument (CGIAR-CSI, 2020; Li, 2003). The derived topographical variables included topographical position index (TPI), terrain ruggedness index (TRI), roughness, aspect and hillshade.

2.2.2.3. Bioclimatic variables. Bioclimatic variables used in this study were obtained from AfriClim (Fick and Hijmans, 2017; Platts et al., 2015) at 1 km spatial resolution. Bioclimatic variables contain derived summaries of rainfall and temperature and describe current (year 2021 - historical data over 1970 – 2000) and future (2055 - simulated means from 2041 to 2070) conditions. Simulated climatic conditions under intermediate CO_2 emissions, set by the International Panel on Climate Change (IPCC) at representative concentration pathway scenario (RCP) 4.5 W/m^2 (Platts et al., 2015) using total radioactive forcing values were used. Twenty-one bioclimatic variables were used for current and future timesteps, comprising of 10 temperature-related variables i.e.; mean annual temperature (*bio1*), mean diurnal range in temperature isothermality (*bio2*), isothermality (*bio3*), temperature seasonality (*bio4*), maximum temperature warmest month (*bio5*), minimum temperature coolest month (*bio6*), annual temperature range (*bio7*), mean temperature warmest quarter (*bio10*), mean temperature coolest quarter (*bio11*) and potential evapotranspiration (*pet*), and 11 precipitation-related variables i.e.; mean annual rainfall (*bio12*), rainfall wettest month (*bio13*), rainfall driest month (*bio14*), rainfall seasonality (*bio15*), rainfall wettest quarter (*bio16*), rainfall driest quarter (*bio17*), annual moisture index (*mi*), moisture index moist quarter (*mimq*), moisture index arid quarter (*miaq*), number of dry months (*dm*) and length of longest dry season (*llds*).

2.2.2.4. Human footprint. The human footprint variable was obtained from the global human footprint (IGHP) data of the last of the wildlife project, version 2 (v2) of 2005 (LWP-2). The dataset was accessed from National Aeronautics and Space Administration's (NASA) socioeconomic data and applications centre' (SEDAC) platform (https://sedac.ciesin.columbia.edu/data/set/wildareas-v2-human-footprint-ighp/data-download), which provides an anthropogenic effect on natural environments across the globe (WCS and CIESIN, 2005). Specifically, the dataset provides the human influence index (HII) normalized based on biomes. It has a 1-kilometre spatial resolution and integrates human population pressure, landuse, infrastructure and human access.

2.2.3. Variable selection

Variables used for model predictions were carefully selected to avoid multicollinearity that could lead to volatility in model performance (Dormann et al., 2013). However, before testing for multicollinearity, the recursive feature elimination (RFE) criteria in the 'caret' package in R was used to provide insight on the minimum number of uncorrelated variables that could yield comparable prediction results (Darst et al., 2018; Makori et al., 2022). To select preferred prediction variables, a two-stage elimination criterion was performed, firstly using variable inflation factor (VIF) and Pearson correlation coefficient. This was meant to reduce multicollinearity amongst predictor variables while establishing orthogonal variables that were most suited for bee pests prediction models (Dormann et al., 2013). Secondly, multiple regression models were utilized to regress each predictor variable against all other variables to detect collinearity while computing VIF for each combination (Plant, 2012). This step was used to select important prediction variables from the pool while iteratively eliminating those with high linear regression coefficients.

The Pearson correlation coefficient was set at $th = 0.7$ ($r \geq 0.7$) as the first threshold for best results. The second threshold was set using the 'vifstep' function in 'usdm' package in R (Naimi and Araújo, 2016). This was used to further assess collinearity among variables from the first step while eliminating collinear ones with more than 10 VIF values. The resultant correlation matrix (Fig. 2) indicated that some predictor variables tagged important for bee pests' distribution were correlated. For instance, start of season time (*start_t*) and rainfall driest quarter (*bio17*) were highly correlated. Hence, 14 important and uncorrelated predictor variables were selected for building the models when fitted with response variables (Araújo et al., 2019).

2.3. Fitting modelling environment

Bee pests' abundance observations were related to selected uncorrelated predictor variables using machine learning random forest (RF) algorithm. The RF ecological niche modelling environment was employed in the 'sdm' package (Naimi and Araújo, 2016) in R software (R Core Team, 2021). Each RF model was set to ten iterations model runs and an ensemble approach was utilized to estimate species-specific mean predictions. Hence, the variations among the predictions were harmonized (Hao et al., 2019). The 'ensemble' function was used in the 'sdm' package to harmonize results of individual bee pests' species prediction replications using true skills statistics (TSS) weighted approach in both the current and future timestamps. A 0.7 TSS cut off was used to select the replicate models that were included in the ensemble (Hao et al., 2019).

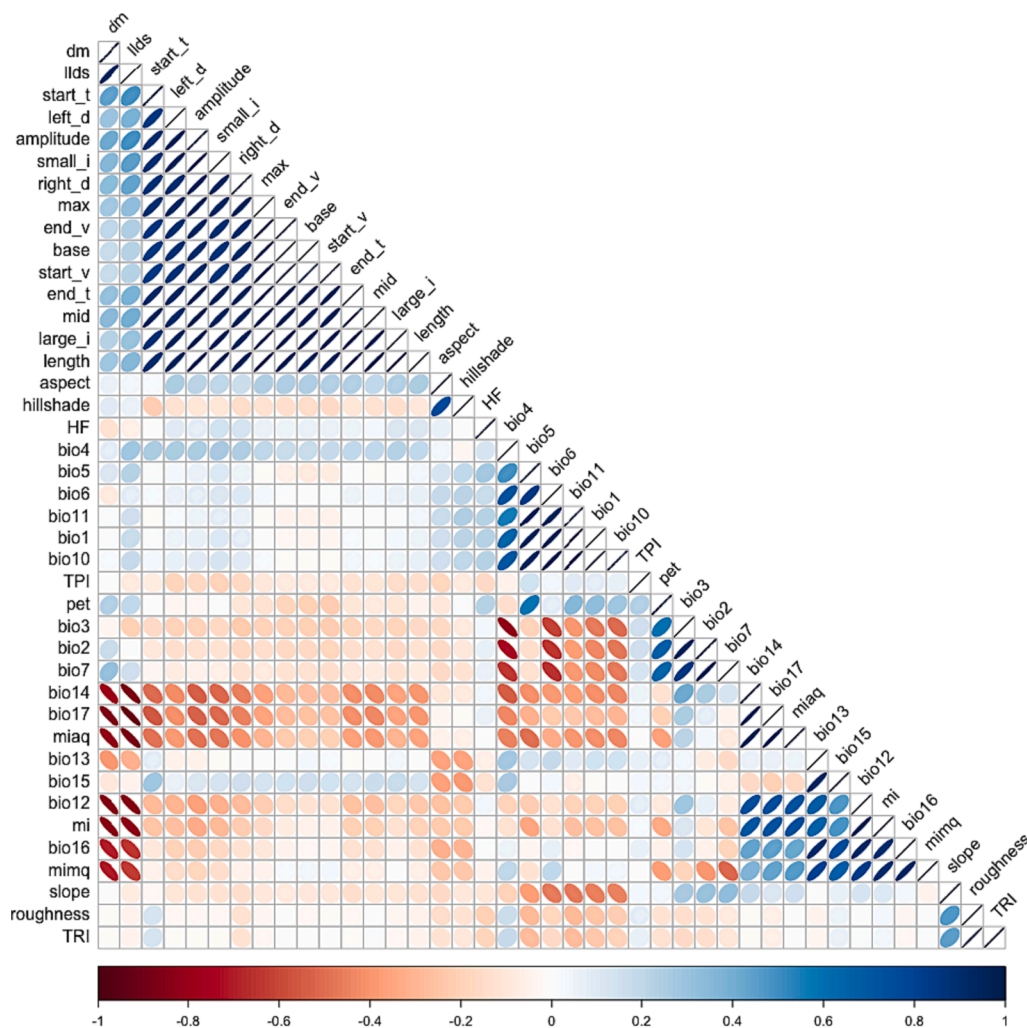Predictions were done under both current and future projections for

Fig. 2. Collinearity matrix indicating correlation interaction of all predictor variables used to predict spatial proliferation of bee pests in Kenya. Red colour denotes negative correlation while blue indicates positive correlation. Darker shades of both red and blue colours indicate higher correlation while size and direction of the circles show extent and nature of correlation. Circles facing the right side indicate positive correlation and vice versa. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

all bee pests' species under similar model settings, with simulated climatic and human footprint scenarios. Since projected vegetation phenology was unavailable and elevation was assumed not to change in the future epoch, their current datasets were also used for future predictions.

2.4. Prediction model validation

The RF models were evaluated using the coefficient of determination (R^2) (Equation 1) that was calculated using the R software (Chicco et al., 2021; Colin Cameron and Windmeijer, 1997; Renaud and Victoria-Feser, 2010; Rights and Sterba, 2020). This established the power of fit between abundance and environmental suitability represented as deviance of fit and the extent which the models could explain the bee pest's abundance distribution. Degree of agreement, sign and significance of the models were used to report suitability and prediction power of each RF model. Prior to the evaluation, data from each model were subjected to covariation linear test and scatter plot reported for each prediction model. Then, model prediction power was ranked using a modified Swets discriminatory power (Makori et al., 2022; Swets, 1988) for each model from fail to excellent (Table 1).

Equation 1: Coefficient of determination (R^2) used to validate the developed bee pests' abundance distribution prediction models.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

where:

 i) SS_{res} is the sum of squares of residuals (unexplained variation)
 ii) SS_{tot} is the total sum of squares (total variation)

3. Results

3.1. Seasonality influence on bee pests' abundance

The Wilcoxon rank sum test with continuity correction performed at 95 percent confidence interval ($p \leq 0.05$) indicated that there was a significant difference ($p = 0.047$) between bee pests' abundance field observations in the wet and dry seasons. Moreover, a visual observation showed high seasonal variability for bee pests (Fig. 3). Generally, there were more bee pests' abundance in the wet than the dry season (Table 2). The average bee pests' abundance in the dry season was 42.25 pest counts with relatively low dispersions around the mean. On the other hand, bee pests' abundance was more (almost five-fold) during the wet season with a mean of 199.75 and higher relative dispersions around the mean compared to the dry season. The *V. destructor* had the highest difference between seasons (at 6.35 times) while *G. mellonella* had the least seasonal difference (at 0.95 times).

3.2. Predictor variable selection and bee pests' abundance predictions

Twenty seven out of forty-one predictor variables used in the collinearity test (Fig. 2) were conflating. Based on their individual interactions and importance they were eliminated from further analysis.

Table 1
Model evaluation ranks used to categorise bee pests' abundance proliferation prediction models developed for *Varroa destructor, Oplostomus haroldi, Galleria mellonella* and *Aethina tumida* in Kenya.

Pearson's correlation coefficient (*r*) value	Rank
0.40 and below	Fail
0.41 to 0.60	Poor
0.61 to 0.70	Fair
0.71 to 0.80	Good
0.81 and above	Excellent

Table 2
Means of *Varroa destructor, Oplostomus haroldi, Galleria mellonella* and *Aethina tumida* abundance data collected during the wet and dry seasons in Kenya. The seasonal difference show difference between seasons.

Bee pest species	Bee pest count		Seasonal difference
	Wet season	Dry season	
V. destructor	476.00	75.00	6.35
O. haroldi	31.00	15.00	2.07
G. mellonella	20.00	21.00	0.95
A. tumida	272.00	58.00	4.69
Mean of total	**199.75**	**42.25**	**4.73**

The selected variables (14) that were non-conflating and ranked high in variable importance scale were used to develop refined bee pests' abundance prediction models. The recursive feature elimination (RFE) model (Fig. 4) used to test the least variable interaction before eliminating the collinear ones further indicated that only 14 predictor variables would yield acceptably high bee pests' abundance predictions.

Bioclimatic and topographical clusters contributed the highest number of variables and logit compared to vegetation phenological variables for all pest-specific models. Furthermore, bioclimatic variables had the highest influence at 37.4% on *V. destructor* prediction models (Fig. 5). Moreover, temperature seasonality and annual temperature range alone had the highest logit at 22.3% on *V. destructor's* prediction. The left derivative and the start of the season value had a substantial single contribution at 14.7% and 14.3%, respectively on predicting bee pests' abundance than any other single variable. Topographical influence on *V. destructor* was 30.1%, while vegetation phenological variables contributed 29.0% logit to this bee pest model. On the other hand, human footprint had the second lowest logit at 3.5% on *V. destructor* prediction.

The bioclimatic variables contributed most logit of 41.1% to *O. haroldi* abundance prediction. Phenological effect was more substantial at 30.6% than topological effect at 23.8%, suggesting that *O. haroldi* was more sensitive to vegetative food substrate than elevational variations. Notably, phenological effect was more pronounced on *G. mellonella* contributing almost half (49.6%) of logit to the prediction model, followed by topology at 29.7%. The bioclimatic effect on the *G. mellonella* was relatively lower at 17.2% compared to the other bee pests. This suggested lower sensitivity of *G. mellonella* to climatic variations and more responsiveness to vegetative food substrate. On the other hand, *A. tumida's* abundance prediction models had a much higher logit for topographical variables at 52.7%, indicating high sensitivity of the bee pest to topographical variations. Also, hillshade and aspect had 27.8 %, more than a quarter logit on the *A. tumida* abundance while bioclimatic and vegetation phonological variables contributed 31.2% and 8.6% logit, respectively. The anthropogenic effect on bee pests' abundance as indicated by human footprint was highest at 7.4% on the *A. tumida* and lowest at 3.5% on the *V. destructor* and *G. mellonella* abundance models.

3.3. Bee pests' abundance prediction validation

The coefficient of determination indicated that the *V. destructor* and *A. tumida* models had a high power of fit between the response and predictor variables, hence regarded as excellent (Table 3). However, the *O. haroldi* model was ranked as good while the *G. mellonella* model was ranked fair. Moreover, although its *p* values at 95% confidence interval were highly significant ($p \leq 0.05$), the *O. haroldi* model had a slightly lower significance score ($p = 2.7e-08$). Furthermore, dispersion of abundance around the line of best fit was higher in the *O. haroldi* and *G. mellonella* models compared to others (Fig. 6). This was in line with observations using the coefficient of determination for the same prediction models, suggesting that *G. mellonella* and *O. haroldi* models were comparatively weaker in fitting bee pests' abundance to environmental suitability than the others.
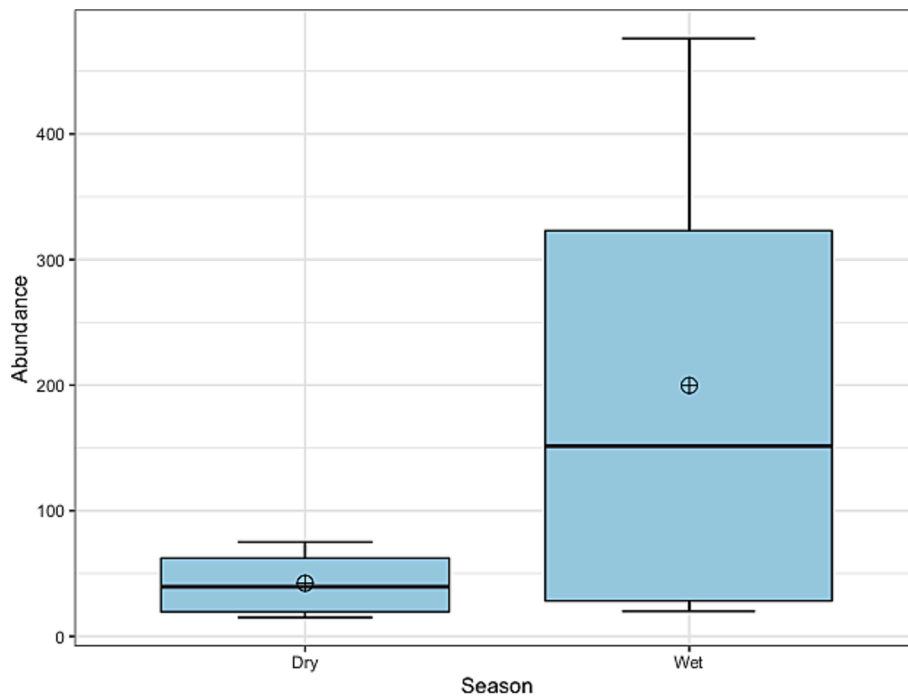
Fig. 3. Boxplot for bee pests' abundance for the dry and wet seasons in Kenya. There was a significant seasonality influence ($p \leq 0.05$) between observations in the two seasons.
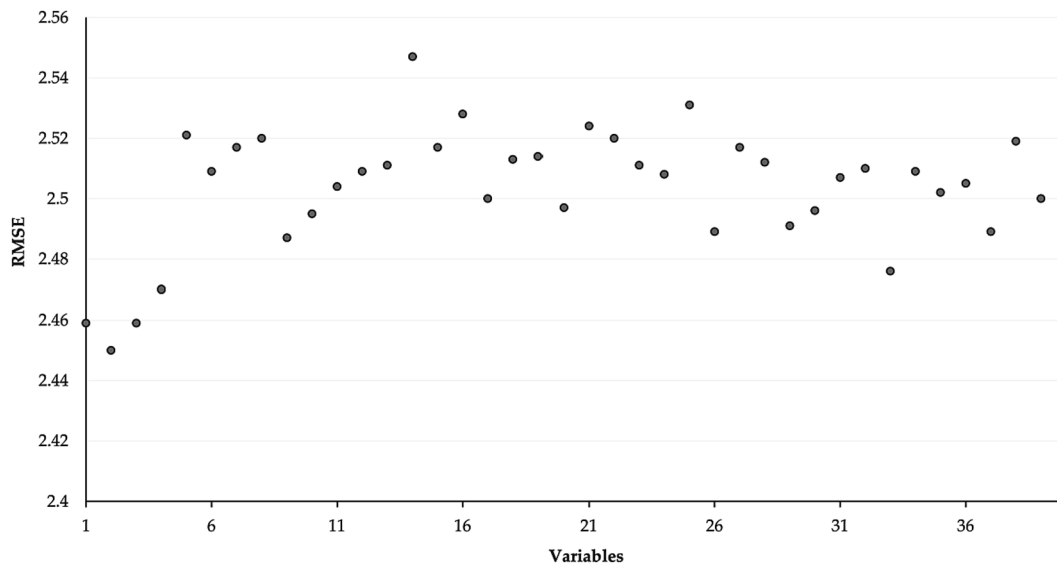

Fig. 4. The recursive feature elimination (RFE) model indicates that 14 independent features (variables) could yield acceptable prediction accuracy based on root mean square error (RMSE).

3.4. Bee pests' proliferation and distribution

The prediction models indicated varied bee pests' abundance in different regions of the study area. The Mt. Kenya region was predicted with a high abundance for *V. destructor* and *A. tumida* and moderate suitability for *G. mellonella*. Moreover, these bee pests were predicted to be highly abundant in the western, Kisii highlands, Lake Victoria, central and coastal regions of Kenya (Fig. 7). These regions have higher precipitation compared to the rest of the study site. Western, Kisii highlands, Mt. Kenya and central regions experience lower temperatures whereas the coastal region experiences higher temperature. On the other hand, the eastern region had varied predicted bee pests' abundances while Tana River, Kajiado and other semi-arid regions were predicted

with low abundance of all bee pests. However, *O. haroldi* had low predicted abundance in most regions of the study area in the current epoch apart from Mwingi in the eastern region. The latter regions are characterized by relatively dry conditions with low precipitation levels, isothermality and high temperatures. These regions exhibit varied agroecological and climatic conditions in line with the predicted abundance. Indeed, rainfall wettest month (*bio13*), rainfall seasonality (*bio15*) and annual temperature range (*bio7*) contributed most logit to the models. On the other hand, high altitude regions were predicted to have a high abundance of all bee pests. Additionally, high precipitation and low temperature were observed in regions with high altitudes. Regions with low altitudes were also characterised by low precipitation levels, isothermality and higher temperature. These regions were
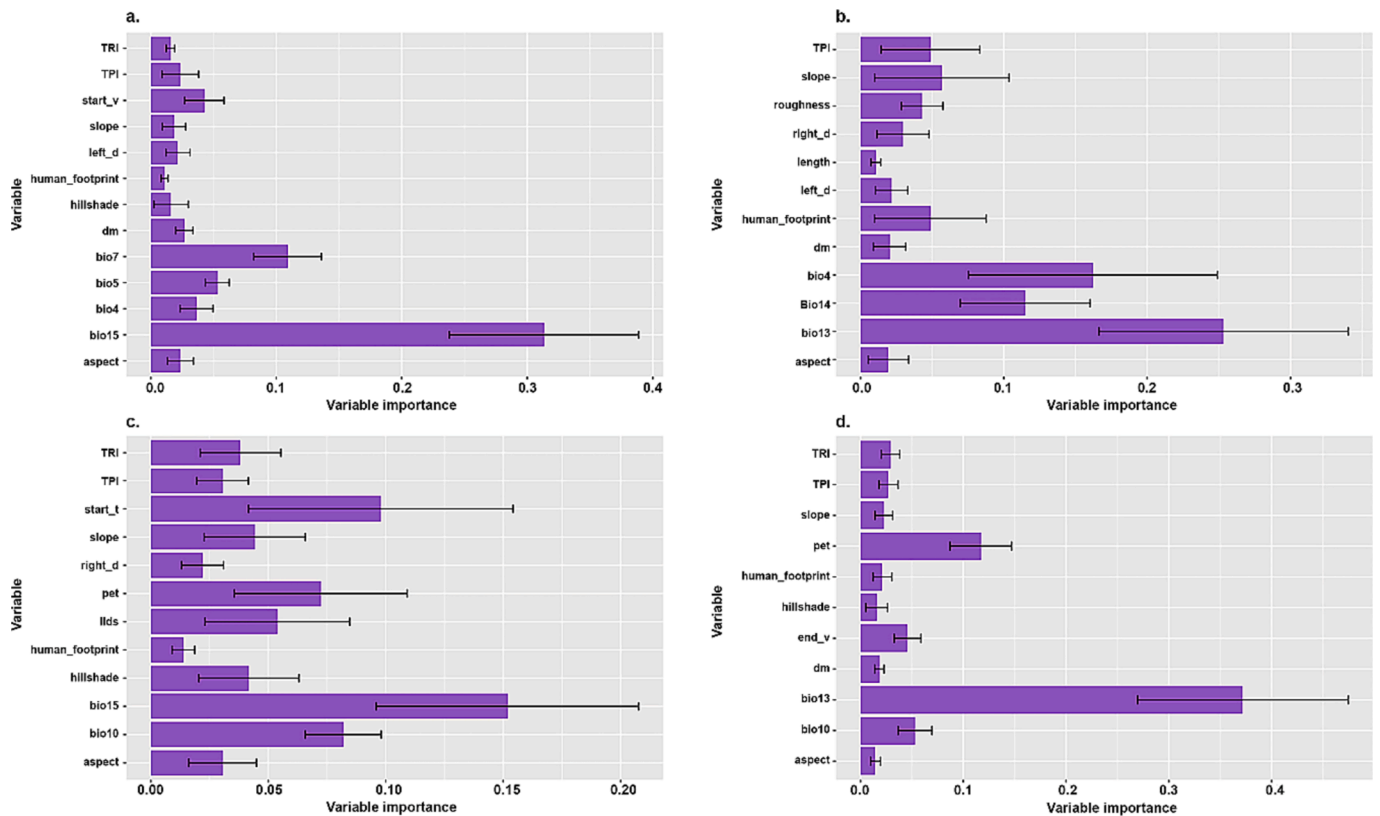
Fig. 5. Interaction of selected non-conflating predictor variables for *Varroa destructor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d) abundance prediction models in Kenya.

Table 3
Coefficient of determination (R^2) and probability scores (*p*-values) of models developed for prediction of abundance of bee pests in Kenya.

Abundance prediction model	R^2	*p*	Rank
Varroa destructor	0.88	6.1e-16	Excellent
Oplostomus haroldi	0.77	2.7e-08	Good
Galleria mellonella	0.67	4.7e-10	Fair
Aethina tumida	0.83	2.5e-11	Excellent

consequently predicted with low bee pests' abundance. They include Narok, Kajiado, Kwale, Tana River, the larger Taita Taveta, Makueni and Kitui counties.

The future abundance predictions indicated a general increase in potential bee pests' risk in most regions of the study area. There was both increment in abundance and spatial proliferation across the study area (Fig. 8). The maps showed that most parts of the study area previously predicted with moderate bee pests' proliferation changed to high abundance in the future epoch. Regions predicted with high proliferation include Mt. Kenya, central, western Kenya, Kisii highlands, Lake Victoria, coast and Mwingi in the eastern region. Notably, *V. destructor* (a) and *A. tumida* (d) had higher proliferation (13,973.15 km^2 and 12,786.62 km^2 respectively) while *O. haroldi* (b) had a moderate proliferation rate (3,076.71 km^2). On the other hand, *G. mellonella* (c) had a higher reduction in proliferation (4,588.16 km^2) than increase in proliferation (2,771.60 km^2) in the future (Table 4). Furthermore, change maps (Fig. 9) indicated that most parts of western, Kisii, Lake Victoria, Mt. Kenya, central and coastal regions had high bee pests' proliferation in the future, while most parts of eastern regions had changed from low to moderate proliferation for *G. mellonella* (c). Moreover, some parts of the study site had reduced proliferation (Table 4 and Fig. 9), which indicated that these sites transformed to less suitability for bee pests. However, areas with reduced bee pests' abundance were less compared

to those with increased proliferation. Bioclimatic data indicated that rainfall seasonality and rainfall wettest quarter increased in the future epoch compared to current conditions.

4. Discussion

There is elevated global interest in beehealth amid increased climate change, disturbances on natural environments and food production systems. As a supplemental source of income and nutritional source, understanding the spatial proliferation of bee threats is vital. Accurate and reliable predictive tools that establish extent and severity of potential risks posed by bee pests' are important in promoting beehealth (Makori et al., 2017). In this study, natural variables and human footprint patterns were utilized to develop bee pests' abundance prediction models. Bee pests' abundance data provided spatially explicit and accurate pest population information for improved predictability as opposed to modelling using occurrence only.

4.1. Seasonality influence on bee pests' abundance

The effect of seasonality on bee pests' abundance was apparent. There was an average of almost five-fold (4.73) more bee pests during the wet season than the dry season. This was more pronounced for *V. destructor* with more than six times (6.35) pests in the wet season compared to the dry season (Table 2). Contrary to this study's findings, previous studies have indicated constant infestation patterns of *V. destructor* across all seasons (Strauss et al., 2013). Other than *G. mellonella*, all bee pests were significantly more in the wet season. Similarly, Torto et. al (2010) reported higher incidences of *A. tumida* during the wet season. Seasonality directly and indirectly affects bee pests' abundance and distribution through their hosts (honeybees), on which they depend for food substrate and feeding on their brood. In addition, the hosts collect nectar and pollen from plants, which are
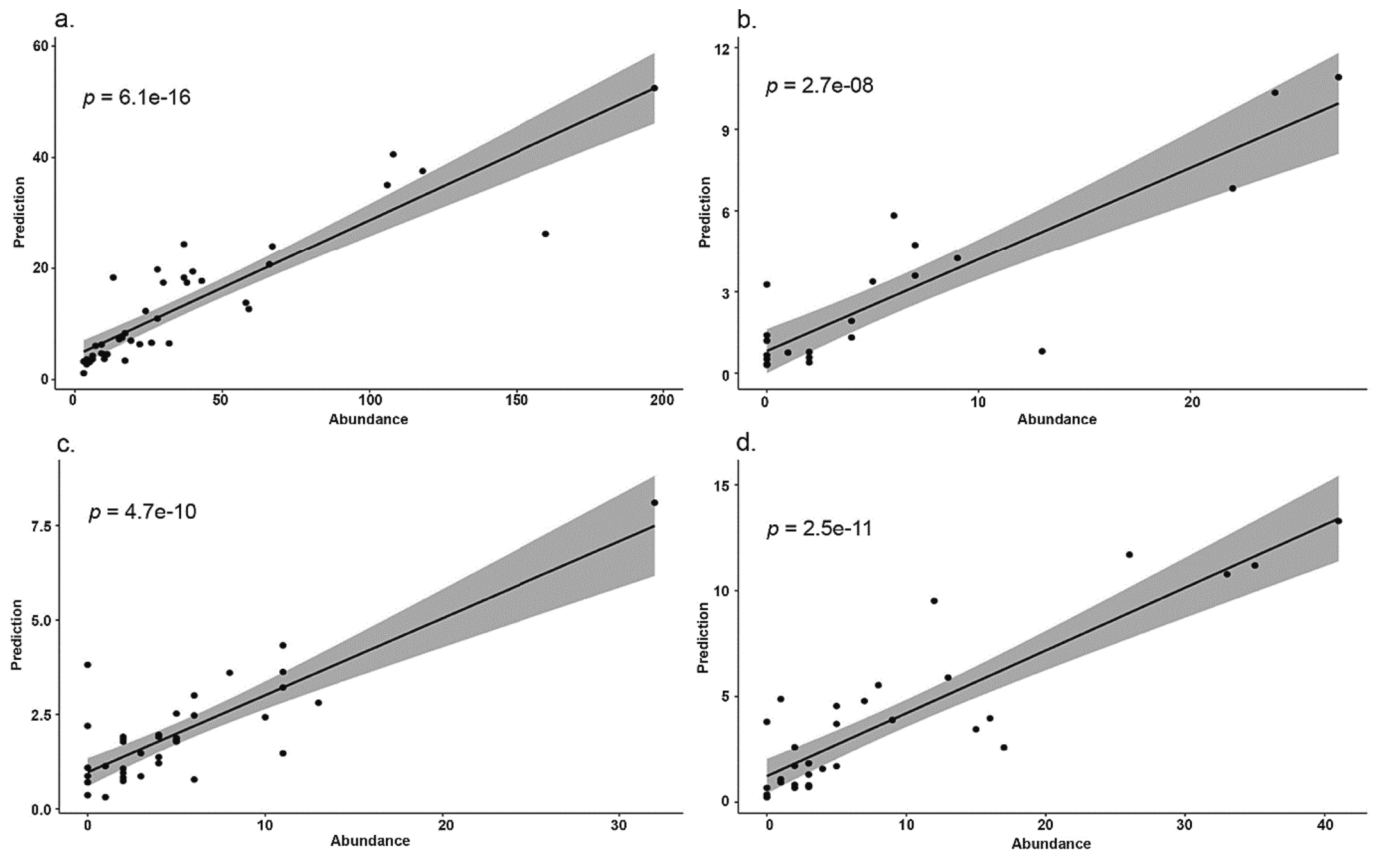
Fig. 6. Scatter plots indicating the distribution of predicted against observed abundance for *Varroa destructor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d).
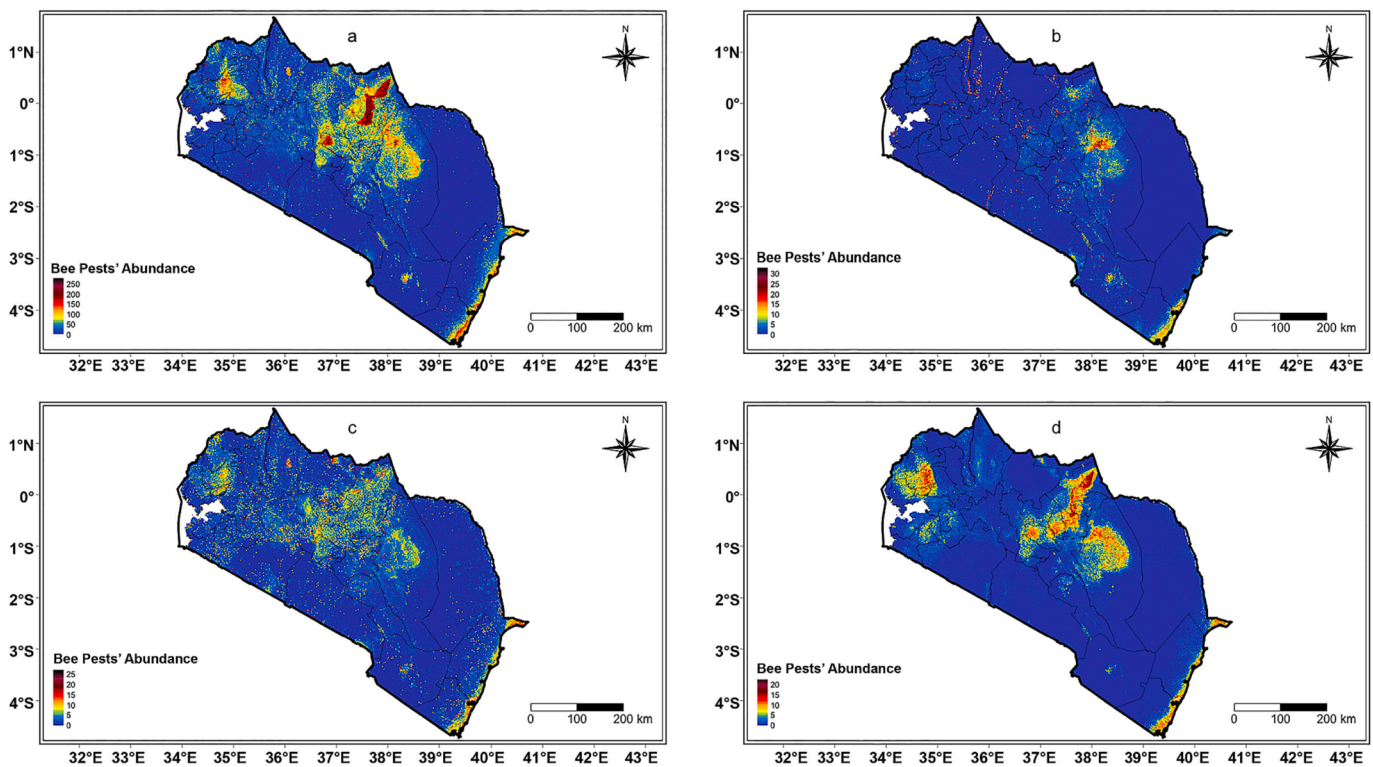

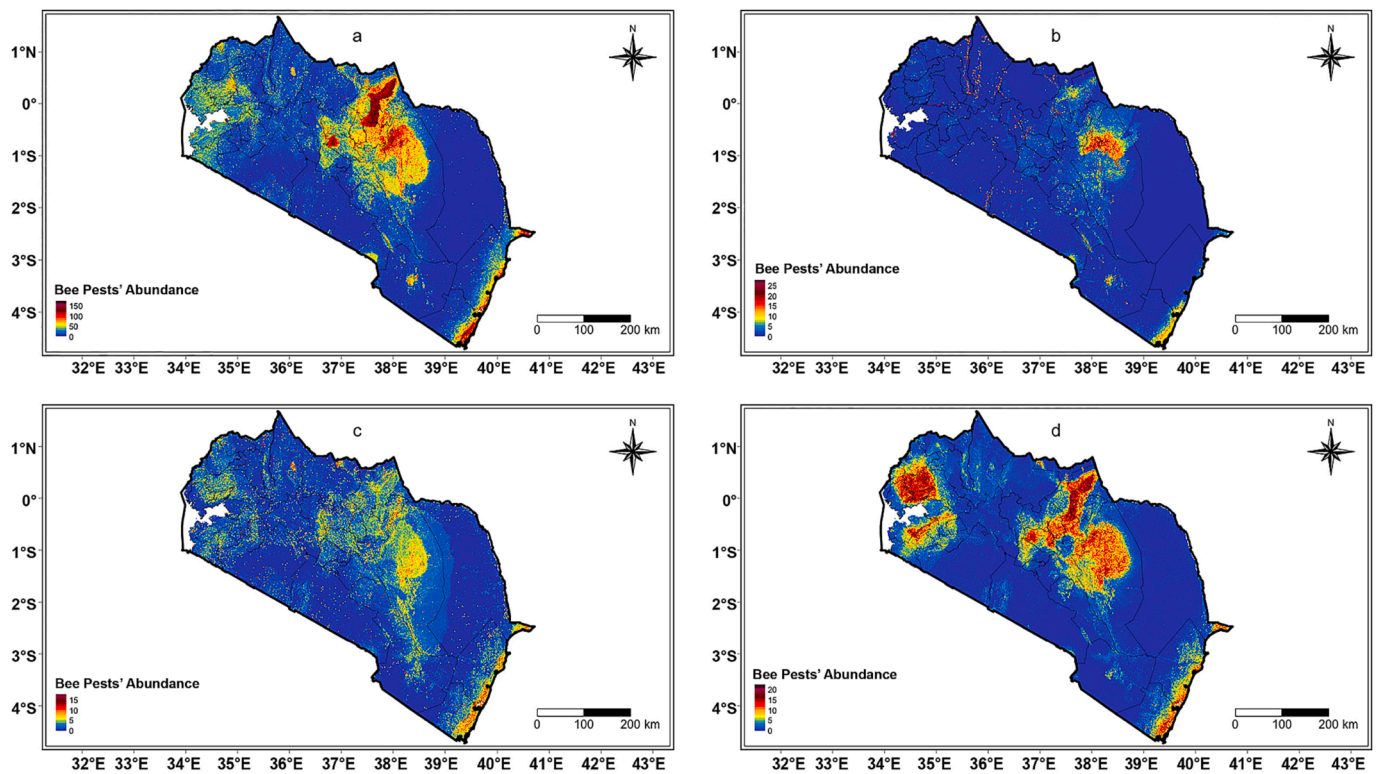**Fig. 7.** Predicted current abundance of *Varroa destructor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d). Blue, yellow and red colours indicate low, moderate and high predicted abundance (pests per apiary), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 8. Predicted future (2055) abundance of *Varroa destructor* (a), *Oplostumus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d). Blue, yellow and red colours indicate low, moderate and high predicted abundance (pests per apiary), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Predicted abundance change for *Varroa destructor, Oplostumus haroldi, Galleria mellonella* and *Aethina tumida* between current and future timestamps represented in area (km²).

Class	Varroa destructor	Oplostomus haroldi	Galleria mellonella	Aethina tumida
	Area (km²)			
Decrease	001,481.43	000,409.36	004,588.16	001,345.22
No change	238,172.53	250,141.03	246,267.34	239,495.27
Increase	013,973.15	003,076.71	002,771.60	012,786.62
Total	**253,627.10**	**253,627.10**	**253,627.10**	**253,627.10**

affected by seasonal variations. Moreover, seasonality affects the ability of honeybees to collect and make food substrate on which the pests' prey. Vegetation phenology indicated a reduction in raw materials (nectar and pollen) in the dry season and vice versa. This is influenced by mean annual rainfall, rainfall seasonality and rainfall wettest month, which are high in the wet season. In addition, both pollen and nectar collected from flowering plants triggered by either onset or end of the rainy season, are available within apiaries for preying by bee pests (Fombong et al., 2012).

4.2. Predictor variable selection and bee pests' abundance prediction

Bee pests proliferate at varying climatic and altitudinal gradients, hence could be found across varying agroecological and agroclimatic regions (Fombong et al., 2012). Their abundance is dependent on their hosts distribution and survival, and various climatic and vegetation conditions. Obligate ectoparasites such as *V. destructor* depend fully on their hosts for survival, hence their success and distribution are dependent on the hosts' resilience and vibrancy. Moreover, generalists' pests

such as *A. tumida* and *G. mellonella* depend on other food substrates as well, hence more susceptible to climatic variations. Predictive models developed in this study established that bioclimatic variables contributed most logit of 17.2% to 41.1% towards the abundance and distribution of bee pests. Additionally, vegetation phenology, topology and human footprint had varied contributions across scale at 3.5% to 52.7%. Moreover, precipitation variables were more predominant than temperature variables for all the studied bee pests.

Rainfall wettest month, rainfall seasonality and annual temperature range contributed most of the information needed for predicting the distribution of bee pests. The models predicted high abundance in regions with high precipitation levels. Moreover, precipitation variation affected distribution and abundance of bee pests across space and time. There was a positive correlation between precipitation and spatial–temporal distribution of all bee pests as indicated by their current increase in abundance and spatial presence in the future epoch (2055). Other studies (Makori et al., 2022, 2017; Mwalusepo et al., 2015) also recorded a positive correlation between precipitation and insects' distribution. Moreover, response variables indicated a significant difference between abundance observations in the wet and dry seasons (Table 2 and Fig. 3). Moreover, Torto et al. (2010) indirectly linked proliferation of *A. tumida* to precipitation via increased forage that triggered host colony growth and increase of brood.

Most predictor variables originally used for prediction of bee pests' abundance in this study were deemed redundant, hence eliminated from final prediction models. The recursive feature elimination (RFE) model was useful in determining the least number of predictor variables that could be used while providing fundamental information necessary for accurate prediction of bee pests' distribution (Fig. 4). In conjunction with collinearity and variable importance elimination criteria, model prediction power was enhanced while minimizing parametrization and volatility. Although vegetation phenological and topological variables were not highly palpable towards bee pests' prediction models and
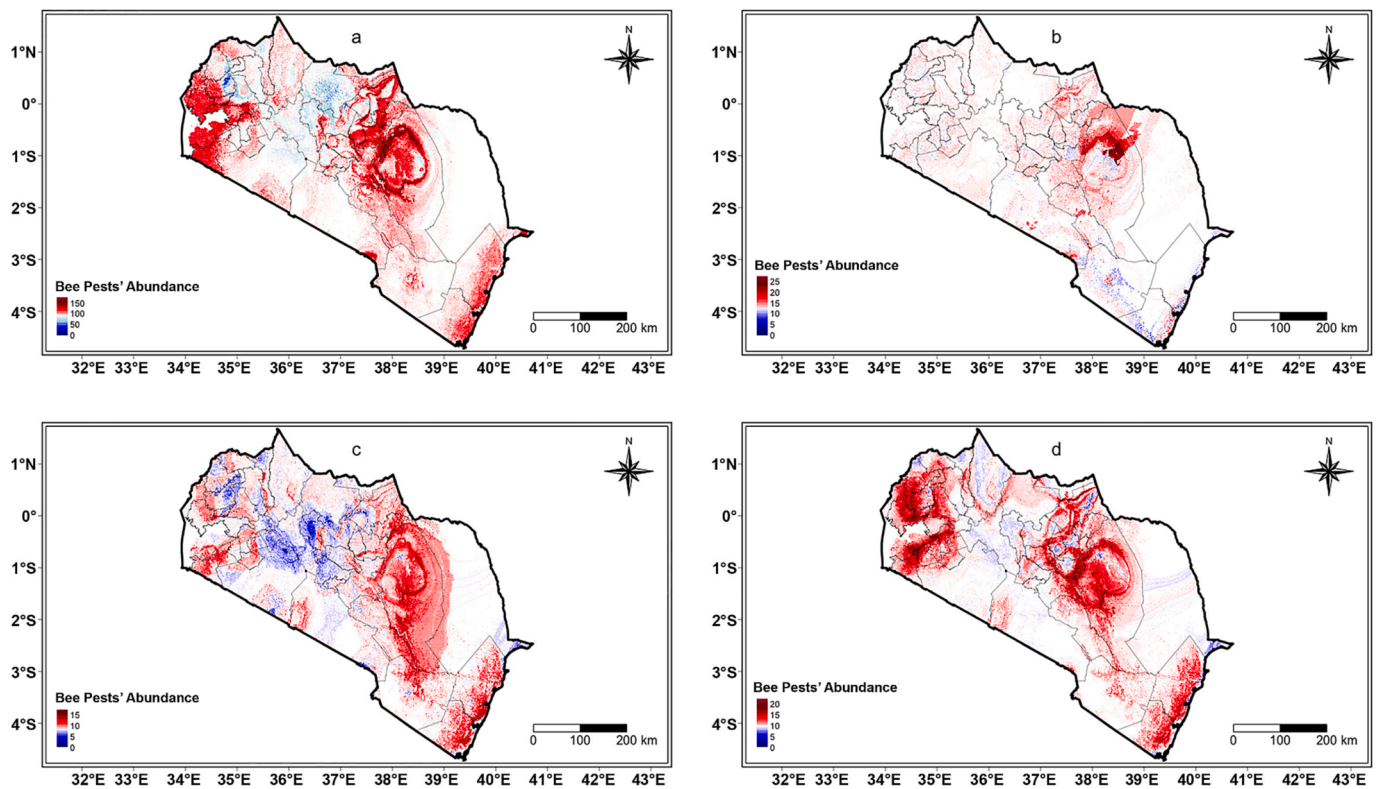
Fig. 9. Change maps for abundance of *Varroa destructor* (a), *Oplostomus haroldi* (b), *Galleria mellonella* (c) and *Aethina tumida* (d). Red and blue colours indicate positive and negative abundance change (pests per apiary), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contributed comparatively lower permutation, they were important in sharpening the models as indicated by the number of variables from these categories tagged (Fig. 5). Moreover, vegetation phenology signifies the availability of both pollen and nectar, which are vital to the pests' hosts in producing food substrates and enhancing colony strength. Both time and value at the start of the season (*start_t* and *start_v* respectively), length of the season (*length*) and right derivative (*right_d*) were among vegetation phenology variables tagged important. All these variables are important ecological indicators of nectar and pollen availability. Right derivative (*right_d*) on the other hand indicates minimum vegetative matter available for bee foraging in between seasons, hence available food substrate to carry both the pests and their hosts through the next season. Mt. Kenya, central, Kisii, Lake Victoria, western and coastal Kenya regions had high values of these predictor variables indicating a positive correlation with bee pests' abundance, which were also high in these regions.

Slope, terrain roughness index (*TRI*) and topographical position index (*TPI*) were among the topographical predictor variables that contributed most logit to prediction models. Combined permutation contribution of topographical predictor variables to bee pests' abundance prediction models ranged from 23.8% to 52.7%. Indeed, most regions predicted with a high abundance of bee pests had high altitude such as Mt. Kenya, Cherangani hills, Kisii highlands and areas around Mt. Elgon. However, coastal Kenya and eastern regions were predicted with a high abundance of bee pests as well. This was in line with previous studies (Kimathi et al., 2020; Makori et al., 2022, 2017) which reported that insects have diverse altitudinal gradients. Although topographical variables were crucial in the prediction of bee pests, their direct impact could not be conclusively deciphered. Their impact could however be derived from altitudinal effect on precipitation, temperature regimes and vegetation phenology. In this regard, different categories of predictor variables could not reliably be used in seclusion to predict distribution of bee pests and should be coupled with other complementary predictor variables.

4.3. Bee pests' abundance prediction validation

Model accuracy is dependent on the selection of appropriate response and non-conflating predictor variables while circumventing redundancy (Araújo et al., 2019). Abundance of bee pests' species was fitted with various predictor variables to determine the best environmental fit. Under the same model settings, two out of four abundance prediction models' performances were ranked excellent (Table 3), indicating a more accurate ordering of specificity and sensitivity. The *O. haroldi* and *G. mellonella* models were ranked good ($R^2 = 0.77$) and fair ($R^2 = 0.67$), respectively with comparatively lower accuracy indicating lower ordering of response and predictor variables. Ordering of observed versus predicted abundance (Fig. 6) indicated low dispersions that pooled closer to the line of the best fit in all models' performance, with slightly higher dispersion on the *O. haroldi* and *G. mellonella* models with notable outlier abundance scores. Hence, these models suggested comparatively higher instability in prediction and thus could not be used in seclusion with high certainty to predict proliferation. Despite notably lower prediction accuracy in the *G. mellonella* model, it had insignificantly low prediction skewness. Consequently, all models developed in this study could be used to order sensitivity and specificity. Nevertheless, rigorous and robust acquisition of substantially observed abundance data is necessary to improve the predictive power of bee pests' abundance prediction models.

4.4. Bee pests' proliferation and distribution

Climate simulation under different CO_2 emission pathways suggests an increase in temperature and localized precipitation intensity in the future (Platts et al., 2015). Hence, the abundance and agility of most beneficial insects such as bees that host some pests will decrease (Torné-

Noguera et al., 2014), reducing their ability to defend themselves against pests. Moreover, prediction models established a positive correlation between levels of precipitation and bee pests' proliferation. Bee pests' abundance increased with an increase in precipitation in some regions in Kenya from moderate to high (Table 4 and Figs. 7, 8 and 9). Furthermore, the Wilcoxon rank sum test with continuity correction indicated significant ($p \leq 0.05$) differences in observed bee pests' abundance between the wet and the dry seasons (Table 2 and Fig. 3), which was more pronounced in *V. destructor* and *A. tumida*. Their prediction models revealed notable differences between predicted bee pests' abundance in current and future epochs. Specifically, there was an increased area suitable for *V. destructor* by a total of 13,973.15 km^2 and a decrease of only 1,481.43 km^2. Moreover, *A. tumida* suitability increased by 12,786.62 km^2 and decreased by only 1,345.22 km^2. The Mt. Kenya, central, Kisii highlands, Lake Victoria, western and coastal Kenya regions, characterized by high precipitation, were most affected. In addition, these regions have been demonstrated to have higher rainfall and temperature seasonality with elevated levels of evapotranspiration in the future. The primal climatic conditions for bee pests could have shifted with interaction of changing bioclimatic variables. In addition, these interactions could affect phenological patterns in the habitats of pests and their hosts. A positive shift of precipitation levels in regions such as Kitui, Narok and Kajiado directly or indirectly triggered an increase of both suitability and abundance of bee pests in the future epoch (Fig. 9). Furthermore, such changes negatively affect bees making them more susceptible to pests' invasion hence increasing their threat levels.

The inclusion of human footprint dataset and vegetation phenology variables improved the predictive power of bee pests' abundance making them more realistic. While human footprint data indicate anthropogenic effect on the prediction models, vegetation phenological variables provided data on growing seasons at grain level, hence availability of food substrates. Moreover, vegetation phenological variables were processed from MODIS EVI at higher spatial and temporal resolutions (250 m) as opposed to bioclimatic data (1,000 m). Therefore, vegetation phenological variables offered the prediction models higher environmental heterogeneity at higher spatial resolution than bioclimatic variables. Hence, vegetation variations enabled the prediction models to identify heterogenous pockets on the landscape which either hinder or limit abundance and proliferation of bee pests. In addition, vegetation phenological variables were derived from 16-day MODIS EVI datasets, which improved the temporal resolution compared to bioclimatic variables that are interpolated over longer periods and large homogenous spatial extents. Therefore, vegetation phenological variables provided prediction models with detailed near real-time and actual information that improved their predictive power, hence more credible and dependable.

5. Conclusions

Amid elevated global interest in climate change and anthropogenic influence on natural environments and agricultural patterns, bee farming has gained relevance, especially in Africa where food and nutritional security are often elusive. However, beehealth is threatened by among others, bee pests, which ravage apiaries. This study developed abundance and distribution prediction models for four bee pests viz *V. destructor, O. haroldi, G. mellonella* and *A. tumida.* Abundance data were integrated with bioclimate, vegetation phenology, topographical and human footprint variables to develop precise and reliable bee pests' abundance prediction models. Two out of four models ranked excellent for bee pests abundancy modelling using the coefficient of determination. Precipitation contributed most logit for the models and seasonal variations proved significantly ($p \leq 0.05$) influential as indicated by the Wilcoxon rank sum test with continuity correction. Furthermore, regions with high rainfall variability and high humidity were predicted with higher threat levels of bee pests. Moreover, threat levels were predicted to increase both spatially and in intensity (abundance) with

climate variations. The prediction models developed in this study could reliably be used to map high risk areas, where management efforts and resources could be employed to curb the spread of bee pests. Therefore, they could provide decision makers with essential tools to assuage spread of bee pests, hence improving beehealth.

CRediT authorship contribution statement

David Masereti Makori: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. **Elfatih M. Abdel-Rahman:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing – review and editing. **John Odindi:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review and editing. **Onisimo Mutanga:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review and editing. **Tobias Landmann:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing. **Henri E.Z. Tonnang:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

References

Adan, M., Abdel-Rahman, E.M., Gachoki, S., Muriithi, B.W., Lattorff, H.M.G., Kerubo, V., Landmann, T., Mohamed, S.A., Tonnang, H.E.Z., Dubois, T., 2021. Use of earth observation satellite data to guide the implementation of integrated pest and pollinator management (IPPM) technologies in an avocado production system.

Remote Sens. Appl. Soc. Environ. 23, 100566 https://doi.org/10.1016/j.rsase.2021.100566.

Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. Sci. Adv. 5, eaat4858. https://doi.org/10.1126/sciadv.aat4858.

Baldridge, E., Harris, D.J., Xiao, X., White, E.P., 2016. An extensive comparison of species-abundance distribution models. PeerJ 4, e2823.

Baumann, M., Ozdogan, M., Richardson, A.D., Radeloff, V.C., 2017. Phenology from Landsat when data is scarce: Using MODIS and Dynamic Time-Warping to combine multi-year Landsat imagery to derive annual phenology curves. Int. J. Appl. Earth Obs. Geoinformation 54, 72–83. https://doi.org/10.1016/j.jag.2016.09.005.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Boncristiani, H., Ellis, J.D., Bustamante, T., Graham, J., Jack, C., Kimmel, C.B., Mortensen, A., Schmehl, D.R., 2021. World Honey Bee Health: The Global Distribution of Western Honey Bee (Apis mellifera L.) Pests and Pathogens. Bee World 98, 2–6. https://doi.org/10.1080/0005772X.2020.1800330.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

CGIAR-CSI, 2020. CGIAR-CSI SRTM – SRTM 90m DEM Digital Elevation Database [WWW Document]. URL http://srtm.csi.cgiar.org/ (accessed 11.8.20).

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623.

Colin Cameron, A., Windmeijer, F.A.G., 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. J. Econom. 77, 329–342. https://doi.org/10.1016/S0304-4076(96)01818-0.

Cord, A.F., Klein, D., Gernandt, D.S., de la Rosa, J.A.P., Dech, S., 2014. Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. J. Biogeogr. 41, 736–748.

Dallas, T.A., Hastings, A., 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. Glob. Ecol. Biogeogr. 27, 1448–1456. https://doi.org/10.1111/geb.12820.

Darst, B.F., Malecki, K.C., Engelman, C.D., 2018. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet. 19, 1–6.

Diao, C., Wang, L., 2014. Development of an invasive species distribution model with fine-resolution remote sensing. Int. J. Appl. Earth Obs. Geoinformation 30, 65–75. https://doi.org/10.1016/j.jag.2014.01.015.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.

Eklundha, L., Jönsson, P., 2017. TIMESAT 3.3 with Seasonal Trend Decomposition and Parallel Processing Software Manual. Lund University.

Endo, T., Watanabe, T., Yamamoto, A., 2015. Confidence interval estimation by bootstrap method for uncertainty quantification using random sampling method. J. Nucl. Sci. Technol. 52, 993–999. https://doi.org/10.1080/00223131.2015.1034216.

Fagerland, M.W., Sandvik, L., 2009. The Wilcoxon–Mann–Whitney test under scrutiny. Stat. Med. 28, 1487–1497. https://doi.org/10.1002/sim.3561.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/joc.5086.

Fombong, A.T., Mumoki, F.N., Muli, E., Masiga, D.K., Arbogast, R.T., Teal, P.E.A., Torto, B., 2012. Occurrence, diversity and pattern of damage of Oplostomus species (Coleoptera: Scarabaeidae), honey bee pests in Kenya. Apidologie 44, 11–20. https://doi.org/10.1007/s13592-012-0149-6.

Hallman, T.A., Robinson, W.D., 2020. Comparing multi- and single-scale species distribution and abundance models built with the boosted regression tree algorithm. Landsc. Ecol. 35, 1161–1174. https://doi.org/10.1007/s10980-020-01007-7.

Hao, T., Elith, J., Guillera-Arroita, G., Lahoz-Monfort, J.J., 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. Divers. Distrib. 25, 839–852. https://doi.org/10.1111/ddi.12892.

Izquierdo-Verdiguier, E., Zurita-Milla, R., 2020. An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. Int. J. Appl. Earth Obs. Geoinformation 88, 102051. https://doi.org/10.1016/j.jag.2020.102051.

Katumo, D.M., Liang, H., Ochola, A.C., Lv, M., Wang, Q.-F., Yang, C.-F., 2022. Pollinator diversity benefits natural and agricultural ecosystems, environmental health, and human welfare. Plant Divers. 44, 429–435. https://doi.org/10.1016/j.pld.2022.01.005.

Kganyago, M., Odindi, J., Adjorlolo, C., Mhangara, P., 2018. Evaluating the capability of Landsat 8 OLI and SPOT 6 for discriminating invasive alien species in the African Savanna landscape. Int. J. Appl. Earth Obs. Geoinformation 67, 10–19. https://doi.org/10.1016/j.jag.2017.12.008.

Kimathi, E., Tonnang, H.E.Z., Subramanian, S., Cressman, K., Abdel-Rahman, E.M., Tesfayohannes, M., Niassy, S., Torto, B., Dubois, T., Tanga, C.M., Kassie, M., Ekesi, S., Mwangi, D., Kelemu, S., 2020. Prediction of breeding regions for the desert locust Schistocerca gregaria in East Africa. Sci. Rep. 10, 11937. https://doi.org/10.1038/s41598-020-68895-2.

Li, J., 2003. Detecting Urban Land-Use and Land-Cover Changes in Mississauga Using Landsat TM Images. J. Environ. Inform. 2, 38–47. https://doi.org/10.3808/jei.200300016.

Makori, D.M., Fombong, A.T., Abdel-Rahman, E.M., Nkoba, K., Ongus, J., Irungu, J., Mosomtai, G., Makau, S., Mutanga, O., Odindi, J., Raina, S., Landmann, T., 2017. Predicting Spatial Distribution of Key Honeybee Pests in Kenya Using Remotely Sensed and Bioclimatic Variables: Key Honeybee Pests Distribution Models. ISPRS Int. J. Geo-Inf. 6, 66. https://doi.org/10.3390/ijgi6030066.

Makori, D.M., Abdel-Rahman, E.M., Ndungu, N., Odindi, J., Mutanga, O., Landmann, T., Tonnang, H.E.Z., Kiatoko, N., 2022. The use of multisource spatial data for determining the proliferation of stingless bees in Kenya. Giscience Remote Sens. 59, 648–669. https://doi.org/10.1080/15481603.2022.2049536.

Muli, E., Patch, H., Frazier, M., Frazier, J., Torto, B., Baumgarten, T., Kilonzo, J., Kimani, J.N., Mumoki, F., Masiga, D., Tumlinson, J., Grozinger, C., 2014. Evaluation of the Distribution and Impacts of Parasites, Pathogens, and Pesticides on Honey Bee (Apis mellifera) Populations in East Africa. PLOS ONE 9, e94459.

Mwalusepo, S., Tonnang, H.E.Z., Massawe, E.S., Okuku, G.O., Khadioli, N., Johansson, T., Calatayud, P.-A., Le Ru, B.P., 2015. Predicting the Impact of Temperature Change on the Future Distribution of Maize Stem Borers and Their Natural Enemies along East African Mountain Gradients Using Phenology Models. PLoS ONE 10, e0130427.

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39, 368–375. https://doi.org/10.1111/ecog.01881.

Parmar, A., Katariya, R., Patel, V., 2019. A Review on Random Forest: An Ensemble Classifier, in: Hemanth, J., Fernando, X., Lafata, P., Baig, Z. (Eds.), International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, Lecture Notes on Data Engineering and Communications Technologies. Springer International Publishing, Cham, pp. 758–763. 10.1007/978-3-030-03146-6_86.

Peterson, A.T., Nakazawa, Y., 2008. Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. Glob. Ecol. Biogeogr. 17, 135–144. https://doi.org/10.1111/j.1466-8238.2007.00347.x.

Pirk, C.W.W., Strauss, U., Yusuf, A.A., Démares, F., Human, H., 2016. Honeybee health in Africa—a review. Apidologie 47, 276–300. https://doi.org/10.1007/s13592-015-0406-6.

Plant, R.E., 2012. Spatial Data Analysis in Ecology and Agriculture Using R. CRC Press.

Platts, P.J., Omeny, P.A., Marchant, R., 2015. AFRICLIM: high-resolution climate projections for ecological applications in Africa. Afr. J. Ecol. 53, 103–108. https://doi.org/10.1111/aje.12180.

QGIS Development Team, 2022. Quantum Geographical Information System [WWW Document]. URL https://www.qgis.org/en/site/ (accessed 6.6.22).

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vigorous Calisthenics.

Raina, S.K., Kioko, E., Zethner, O., Wren, S., 2011. Forest Habitat Conservation in Africa Using Commercially Important Insects. Annu. Rev. Entomol. 56, 465–485. https://doi.org/10.1146/annurev-ento-120709-144805.

Renaud, O., Victoria-Feser, M.-P., 2010. A robust coefficient of determination for regression. J. Stat. Plan. Inference 140, 1852–1862. https://doi.org/10.1016/j.jspi.2010.01.008.

Rights, J.D., Sterba, S.K., 2020. New Recommendations on the Use of R-Squared Differences in Multilevel Model Comparisons. Multivar. Behav. Res. 55, 568–599. https://doi.org/10.1080/00273171.2019.1660605.

Schürmann, A., Kleemann, J., Fürst, C., Teucher, M., 2020. Assessing the relationship between land tenure issues and land cover changes around the Arabuko Sokoke Forest in Kenya. Land Use Policy 95, 104625.

Speiser, J.L., Miller, M.E., Tooze, J., Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst. Appl. 134, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028.

Strauss, U., Human, H., Gauthier, L., Crewe, R.M., Dietemann, V., Pirk, C.W.W., 2013. Seasonal prevalence of pathogens and parasites in the savannah honeybee (Apis mellifera scutellata). J. Invertebr. Pathol. 114, 45–52. https://doi.org/10.1016/j.jip.2013.05.003.

Strebel, N., Kéry, M., Guélat, J., Sattler, T., 2022. Spatiotemporal modelling of abundance from multiple data sources in an integrated spatial distribution model. J. Biogeogr. 49, 563–575. https://doi.org/10.1111/jbi.14335.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293. https://doi.org/10.1126/science.3287615.

Torné-Noguera, A., Rodrigo, A., Arnan, X., Osorio, S., Barril-Graells, H., da Rocha-Filho, L.C., Bosch, J., 2014. Determinants of Spatial Distribution in a Bee Community: Nesting Resources, Flower Resources, and Body Size. PLoS ONE 9. https://doi.org/10.1371/journal.pone.0097255.

Torto, B., Fombong, A.T., Mutyambai, D.M., Muli, E., Arbogast, R.T., Teal, P.E.A., 2010. Aethina tumida (Coleoptera: Nitidulidae) and Oplostomus haroldi (Coleoptera: Scarabaeidae): Occurrence in Kenya, Distribution within Honey Bee Colonies, and Responses to Host Odors. Ann. Entomol. Soc. Am. 103, 389–396. https://doi.org/10.1603/AN09136.

Uusitalo, R., Siljander, M., Culverwell, C.L., Mutai, N.C., Forbes, K.M., Vapalahti, O., Pellikka, P.K.E., 2019. Predictive mapping of mosquito distribution based on environmental and anthropogenic factors in Taita Hills, Kenya. Int. J. Appl. Earth Obs. Geoinformation 76, 84–92. https://doi.org/10.1016/j.jag.2018.11.004.

Waldock, C., Stuart-Smith, R.D., Albouy, C., Cheung, W.W.L., Edgar, G.J., Mouillot, D., Tjiputra, J., Pellissier, L., 2022. A quantitative review of abundance-based species distribution models. Ecography 2022. https://doi.org/10.1111/ecog.05694.

Warui, M.W., Gikungu, M., Bosselmann, A.S., Hansted, L., 2018. Pollination of Acacia woodlands and honey production by honey bees in Kitui, Kenya. Future Food J. Food Agric. Soc. 6, 40–50.

WCS, CIESIN, 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (IGHP). 10.7927/H4GF0RFQ.

Winfree, R., Aguilar, R., Vázquez, D.P., LeBuhn, G., Aizen, M.A., 2009. A meta-analysis of bees' responses to anthropogenic disturbance. Ecology 90, 2068–2076. https://doi.org/10.1890/08-1245.1.

Wulder, M.A., Roy, D.P., Radeloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M., Healey, S., Zhu, Z., Scambos, T.A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C.J., Masek, J.G., Hermosilla, T., White, J.C., Belward, A.S., Schaaf, C., Woodcock, C.E., Huntington, J.L., Lymburner, L., Hostert, P., Gao, F., Lyapustin, A.,

Pekel, J.-F., Strobl, P., Cook, B.D., 2022. Fifty years of Landsat science and impacts. Remote Sens. Environ. 280, 113195 https://doi.org/10.1016/j.rse.2022.113195.

Zhang, X., Liao, C., Li, J., Sun, Q., 2013. Fractional vegetation cover estimation in arid and semi-arid environments using HJ-1 satellite hyperspectral data. Int. J. Appl. Earth Obs. Geoinformation 21, 506–512. https://doi.org/10.1016/j.jag.2012.07.003.